2018

Opinion Broadcasting Model

Chenchao Chen
University of Pennsylvania, chencc.math.pku@gmail.com

Opinion Broadcasting Model

Abstract
Stochastic approximation methods have been widely used in random processes with reinforcement. Applications include root approximation algorithm, interacting urn models, and continuous reinforced processes. Emphasis is on the establishment of a so-called opinion broadcasting model, with the proof of an asymptotic result for such process using the idea of some stochastic approximation method.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Mathematics

First Advisor
Robin Pemantle

Subject Categories
Mathematics

OPINION BROADCASTING MODEL

Chenchao Chen

A DISSERTATION

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

____
Robin Pemantle, Professor of Mathematics

Graduate Group Chairperson

____
Wolfgang Ziller, Professor of Mathematics

Dissertation Committee:
Robin Pemantle, Professor of Mathematics
J. Michael Steele, Professor of Statistics
Greta Panova, Assistant Professor of Mathematics

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Robin Pemantle for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. J. Michael Steele and Prof. Greta Panova, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I thank my fellow Ph.D colleague Konstantinos Karatapanis for the stimulating discussions, for very careful review of my draft papers, and for all the valuable ideas and suggestions on my project. Also I thank my friends in math department for the time I spent here.

Last but not the least, I would like to thank my parents for supporting me spiritually throughout writing this thesis and my life in general.

ABSTRACT

OPINION BROADCASTING MODEL

Chenchao Chen

Robin Pemantle

Stochastic approximation methods have been widely used in random processes with reinforcement. Applications include root approximation algorithm, interacting urn models and continuous reinforced processes. Emphasis is on the establishment of a so-called opinion broadcasting model, with the proof of an asymptotic result for such process using the idea of some stochastic approximation method.

Contents

Chapter 1

Introduction

The Random process with Reinforcement has grown substantially within the past twenty years. In some sense, it is still a collection of disjoint techniques. The few difficult open problems that have been solved have not led to broad theoretical advances. On the other hand, some nontrivial mathematics is being put to use in a fairly coherent way by communities of social and biological scientists. As a result, several useful techniques are introduced, among which, the stochastic approximation / dynamical system approach and the multitype branching process approach could be called theories which contain their own terminology, constructions, fundamental results, compelling open problems and so forth. However, the later one pioneered by Athreya and Karlin has been taken pretty much to completion by the work of S. Janson.

There is one more area that seems fertile if not yet coherent, namely reinforcement in continuous time and space. Continuous reinforcement processes are to reinforced random walks what Brownian motion is to simple random walk, that is to say, there are new layers of complexity. There are several self-interacting diffusions and more general continuous-time processes that open up mathematics of some depth and practical relevance.

Opinion propagation model is a particular example fit into the catagory of random process with reinforcement, which originally comes from one of Elchanan Mossel's former students. This resulting paper will study a special case of this model using stochastic approximation / dynamic system via martingale methods. There are some potential open questions relavant to this model which either requires more delicate analysis or encounters new form of stochastic approximation.

The organization of the rest of the paper is as follows. Chapter 2 provides an overview of stochastic approximation methods, dynamic system approach and their applications. Chapter 3 introduces the opinion broadcasting model and its motivation, listing the main result and further conjectures. Chapter 4 is devoted to the proofs for the main result.

Chapter 2

Background

The purpose of this chapter is to introduce the stochastic approximation process, dynamic system approach and several of their applications. Because of the way research has developed, the existing theoretical results are very much tailored to speific applications and are not easily discussed abstractly. We start with the root approximation model in chapter 2.1 where the stochastic approximation method first introduced. We further look at its applications in urn models in chapter 2.2. Finally, chapter 2.3 is about the dynamical systems and their stochastic counterparts.

2.1 Root approximation model

In [RM51], the stochastic approximation process was first introduced by Herbert Robbins and Sutton Monro. They used this to approximate the root of an unknown function in the setting where evaluation queries may be made but the answers are noisy.

To be more specific, let $M(x)$ be a given function and α a given constant such that the equation

$$M(x) = \alpha \tag{2.1.1}$$

has unique root $x = \theta$. There are many methods for determining the value of θ by successive approximation (root-finding algorithm). With any such method we begin by choosing one or more values $x_1, \cdots, x_r$ more or less arbitrarily, and then successively obtain new values x_n as certain functions of the previously obtained $x_1, \cdots, x_{n-1}$, the values $M(x_1), \cdots, M(x_{n-1})$ and possibly those of the derivatives $M'(x_1), \cdots, M'(x_{n-1})$, etc. If

$$\lim_{n\to\infty} x_n = \theta, \tag{2.1.2}$$

irrespective of the arbitrary initial values $x_1, \cdots, x_r$, then the method is effective for the particular function $M(x)$ and value α.

The stochastic generalization of the above problem in which the nature of the function $M(x)$ is unknown assumes that to each value x corresponds a random variable $Y = Y(x)$ with distribution $P(Y(x) \leq y) = H(y|x)$ such that

$$M(x) = \int_{-\infty}^{\infty} y dH(y|x)$$

is the expected value of Y for the given x. Neither $H(y|x)$ nor $M(x)$ is known here, but it is assumed that equation (2.1.1) has a unique root θ, and it is desired to estimate θ by making successive observations on Y at levels $x_1, x_2, \cdots$ determined sequentially in accordance with some definite experimental procedure. [RM51] gives a particular procedure for estimating θ which is consistent under certain restrictions on the nature of $H(y|x)$ where the consistency is in the sense that (2.1.2) holds in probability irrespective of any arbitrary initial values $x_1, \cdots, x_r$.

In the proofs, they defined a fixed sequence $\{a_n\}$ of positive constants such that

$$0 < \sum_n a_n^2 = A < \infty.$$

Furthermore, they defined a Markov chain $\{x_n\}$ by taking x_1 to be an arbitrary constant and defining

$$x_{n+1} - x_n = a_n(\alpha - y_n), \tag{2.1.3}$$

where y_n is a random variable such that

$$P(y_n \leq y|x_n) = H(y|x_n).$$

A natural candidate for a_n can be $a_n = \frac{1}{n}$, then (2.1.3) becomes

$$x_{n+1} - x_n = \frac{1}{n}(\alpha - y_n) \tag{2.1.4}$$

where $y_n \in \mathcal{F}_n$. (2.1.4) is an example of a stochastic approximation process.

More generally, let $\{\mathbf{X}_n : n \geq 0\}$ be a stochastic process in the euclidean space $\mathcal{R}^d$ and adapted to a filtration $\{\mathcal{F}_n\}$. Suppose that $\mathbf{X}_n$ satisfies

$$\mathbf{X}_{n+1} - \mathbf{X}_n = \frac{1}{n}(F(\mathbf{X}_n) + \xi_{n+1} + \mathbf{R}_n) \tag{2.1.5}$$

where F is a vector field on $\mathcal{R}^d$, $E(\xi_{n+1}|\mathcal{F}_n) = 0$ and the remainder terms $\mathbf{R}_n \in \mathcal{F}_n$ go to zero and satisfy $\sum_{n=1}^{\infty} n^{-1}|\mathbf{R}_n| < \infty$ almost surely. Such a process is known as a stochastic approximation process.

2.2 Urn models

The original Pólya urn model which first appeared in [EP23] has an urn that begins with one red ball and one black ball. At each time step, a ball is chosen at random and put back in the urn along with one extra ball of the color drawn, this process being repeated infinitely many times. We construct this recursively: let $R_0 = a$ and $B_0 = b$ for some constants $a, b > 0$; for $n \geq 1$, let $R_{n+1} = R_n + \mathbf{1}_{U_{n+1} \leq X_n}$ and $B_{n+1} = B_n + \mathbf{1}_{U_{n+1} > X_n}$, where $X_n := R_n/(R_n + B_n)$. We interpret R_n as the number of red balls in the urn at time n and B_n as the number of black balls at time n. Uniform drawing corresponds to drawing a red ball with probability X_n independent of the past; this probability is generated by oursource of randomness via the random variable U_{n+1}, with the event $\{U_{n+1} \leq X_n\}$ being the event of drawing a red ball at step n.

Later, Pólya's urn has been generalized by taking the number of colors to be any integer $k \geq 2$. The number of balls of color j at time n will be denoted R_{nj}. Secondly, fix real numbers $\{A_{ij} : 1 \leq i, j \leq k\}$ satisfying $A_{ij} \geq -\delta_{ij}$ where δ_{ij} is the Kronecker delta function. When a ball of color i is drawn, it is replaced in the urn along with A_{ij} balls of color j for $1 \leq j \leq k$. The reason to allow $A_{ii} \in [-1, 0]$ is that we may think of not replacing (or not entirely replacing) the ball that is drawn. Formally, the evolution of the vector $\mathbf{R}_n$ is defined by letting $\mathbf{X}_n := \mathbf{R}_n / \sum_{j=1}^{k} R_{nj}$ and setting $R_{n+1,j} = R_{nj} + A_{ij}$ for the unique i satisfying $\sum_{t<i} X_{nt} < U_{n+1} \leq \sum_{t \leq i} X_{nt}$. This guarantees that $R_{n+1,j} = R_{nj} + A_{ij}$ with probability X_{ni} for each i. We call this model generalized Pólya urn scheme (GPU) where the reinforcement is A_{ij}.

To relate between stochastic approximations and urn processes, we denote by Q_n, the probability distributions governing the color of the next ball chosen which are typically defined to depend on the content vector $\mathbf{R}_n$ only via its normalization $\mathbf{X}_n$. If b new balls are added to N existing balls, the resulting increment $X_{n+1} - X_n$ is exactly $\frac{b}{b+N}(\mathbf{Y}_n - \mathbf{X}_n)$ where $\mathbf{Y}_n$ is the normalizedvector of added balls. Since b is of constant order and N is of order n, the mean increment is

$$E(\mathbf{X}_{n+1} - \mathbf{X}_n | \mathcal{F}_n) = \frac{1}{n}(F((\mathbf{X}_n) + O(n^{-1}))$$

where $F(\mathbf{X}_n) = b \cdot (E(\mathbf{Y}_n | \mathcal{F}_n) - \mathbf{X}_n)$. Defining ξ_{n+1} to be the martingale increment $\mathbf{X}_{n+1} - E(\mathbf{X}_{n+1} | \mathcal{F}_n)$ recovers (2.1.5).

Results like the convergence property of $\mathbf{X}_n$ to stable equilibria [HLS80] and nonconvergence to unstable equilibria [Pem88] are based on this stochastic approximation form of GPU.

2.3 Dynamic systems

In terms of [Ben99], Benaïm and collaborators have formulated an approach to stochastic approximations based on notions of stability for the approximating ODE. Here we briefly describe the dynamical system approach.

For processes in any dimension obeying the stochastic approximation equation (2.1.5), there are two natural heuristics. Sending the noise and remainder terms to zero yields a difference equation $\mathbf{X}_{n+1} - \mathbf{X}_n = n^{-1}F(\mathbf{X}_n)$ and approximating $\sum_{k=1}^{n} k^{-1}$ by continous variable $\log t$ yields the differential equation

$$\frac{d\mathbf{X}}{dt} = F(\mathbf{X}). \tag{2.3.1}$$

The first heuristic is that trajectories of the stochastic approximation $\{\mathbf{X}_n\}$ should approximate trajectories of the ODE $\{\mathbf{X}(t)\}$. The second is that stable trajectories of the ODE should show up in the stochastic system, but unstable trajectories should not.

The whole rigorous theory hugely relied on some pure topological concepts which I will omit here. Instead, I will include those results related to probabilistic analysis from [Ben99].

First of all, we need an important notion, introduced by Benaïm and Hirsch

[BH96], is the asymptotic pseudotrajectory.

Definition 2.3.1 (asymptotic pseudotrajectories)**.** Let $(t, x) \mapsto \Phi_t(x)$ be a flow on a metric space M. For a continuous trajectory $X : \mathcal{R}^+ \to M$, let

$$d_{\Phi,t,T} := \sup_{0 \le h \le T} d(X(t+h), \Phi_h(X(t)))$$

denote the greatest divergence over the time interval $[t, t+T]$ between X and the flow Φ started from $X(t)$. The trajectory X is an asymptotic pseudotrajectory for Φ if

$$\lim_{t \to \infty} d_{\Phi,t,T}(X) = 0$$

for all $T > 0$.

Turning to the first convergence heuristic, from [Ben99] Proposition 4.4 and Theorem 7.3, we have:

Theorem 2.3.2 (stochastic approximations are asymptotic pseudotrajectories)**.** *Let $\{\mathbf{X}_n\}$ be a stochastic approximation process, that is, a process satisfying* (2.1.5)*, and assume F is Lipschitz. Let $\{\mathbf{X}(t) := \mathbf{X}_n + (t-n)(\mathbf{X}_{n+1} - \mathbf{X}_n)$ for $n \le t < n+1\}$ linearly interpolate $\mathbf{X}$ at nonintegral times. Assume bounded noise: $|\xi_n| \le K$. Then $\{\mathbf{X}(t)\}$ is almost surely an asymptotic pseudotrajectory for the flow Φ of integral curves of F.*

Theorem 2.3.3 (convergence to an attractor)**.** *Let A be an attractor for the flow associated to the Lipschitz vector field F, the mean vector field for a stochastic* *approximation* $\mathbf{X} := \{\mathbf{X}_n\}$*. Then either* (i) *there is a* t *for which* $\{\mathbf{X}_{t+s} : s \geq 0\}$ *almost surely avoids some neighborhood of* A *or* (ii) *there is a positive probability that* $L(\mathbf{X}) \subseteq A$*.*

For the nonconvergence heuristic, most known results are proved under linear instability. This is a stronger hypothesis than topological instability, requiring that at least one eigenvalue of dF have strictly positive real part. From [Ben99] Theorem 9.1, we have:

Theorem 2.3.4 (nonconvergence under linear instability)**.** *Let* $\{\mathbf{X}_n\}$ *be a stochastic approximation process on a compact manifold* M *with bounded noise* $|\xi_n| \leq K$ *for all* n *and* C^2 *vector field* F*. Let* Γ *be a linearly unstable equilibrium or periodic orbit for the flow induced by* F*. Then*

$$P(\lim_{n\to\infty} d(\mathbf{X}_n, \Gamma) = 0) = 0.$$

Chapter 3

Opinion propagation model

The opinion propagation model is a general model on any graph that originally comes from one of Elchanan Mossel's former students. Let G be a finite connected graph; the nodes represent people and the edges represent that they interact. Choose an integer $k \geq 2$, the set $[k]$ is interpreted as a set of k different opinions, say possible names for a baby, on which it is desired to form a consensus. The states of the system are maps $\xi : V(G) \to 2^{[k]}$, in other words every person subscribes to some subset of all possible opinions. Maps for which $\xi(v)$ is empty set for some v are not allowed. The problem is to study after exchanging information for how many times (or how long in the continuous version) will the state terminates in consensus. in order words after how much time of discussion will every person agrees of a certain name for the baby.

The Markovian evolution is as follows. At rate 1 each person v chooses a random one of their opinions, and independently chooses a uniform neighbor w. Suppose $j \in [k]$ and $w \in V(G)$ are chosen. If $j \notin \xi(w)$ then the new state $\xi^{'}$ agrees with ξ away from w but has $\xi^{'}(w) = \xi(w) \cup \{j\}$. If already $j \in \xi(w)$ then $\xi^{'}$ agrees with ξ except at the two sites v and w, resetting $\xi^{'}(v) = \xi^{'}(w) = \{j\}$. The interpretation is: if you hear a new name you add it to your list; if you hear one you already have in mind, then immediately you and the person you heard it from coordinate on the new name and forget all other names.

3.1 Main results

There is only one thing known for this model:

Theorem 3.1.1. *If $G = K_N$, the complete graph on N vertices, and $k = 2$ (two possible opinions), then starting from any configuration, the state terminates in consensus in $\Theta(N \log N)$ steps.*

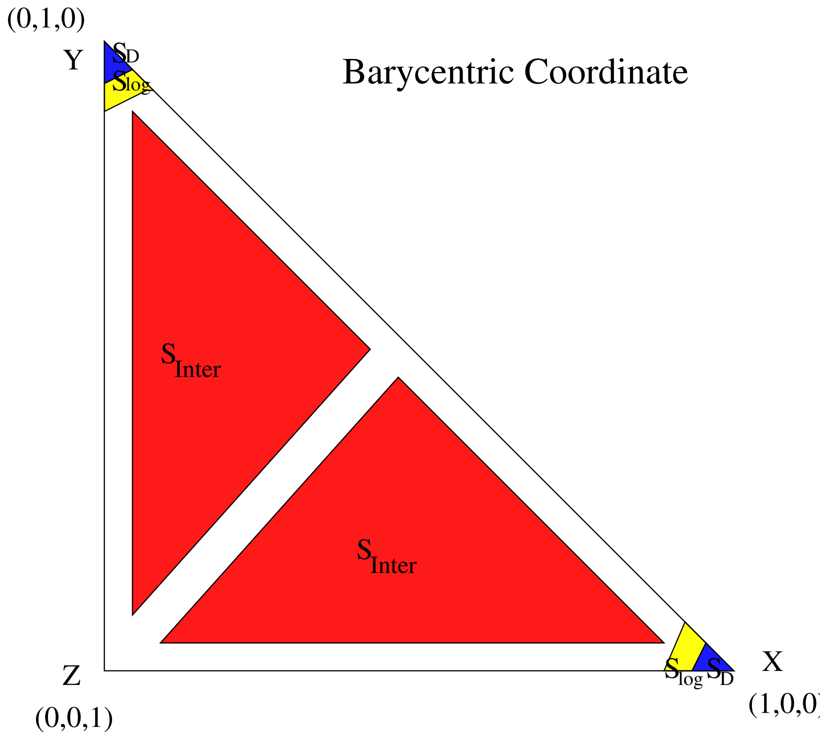
In the rest of the paper, we will improve the result of the above theorem by replacing $\Theta(\cdot)$ with deterministic coefficients for the leading terms. Before listing the main result, we need some notations, basic constructions of them, and some definitions as preparation.

When $k = 2$, we denote the two opinions as opinion A and opinion B. Hence at any time there are at most three types of opinions: $\{A\}$, $\{B\}$, and $\{A, B\}$. We denote X_t^N, Y_t^N and Z_t^N being the proportion of people holding opinion $\{A\}$, $\{B\}$, and $\{A, B\}$ respectively at time t. Obviously we have $X_t^N + Y_t^N + Z_t^N = 1$ for all $t \geq 0$.

Next, we give the following definitions with respect to the area spanned by $(X, Y, Z)_t$'s:

Definition 3.1.2. We denote $\mathcal{T} := \{(X, Y, Z) : 0 \leq X, Y, Z \leq 1, X + Y + Z = 1\}$, and we specify the subset our initial configuration locates as $S_{\text{Inter}} := \{(X, Y, Z) : X, Y, Z \geq 2\epsilon_0, |X - Y| \geq 2\epsilon_0, X + Y + Z = 1\}$ for arbitrary small constant $\epsilon_0 > 0$. Furthermore, we define $S_{\log -dist} := \{(X, Y, Z) : Y^2 + Z^2 \leq \log^{-2} N, Y^2 + Z^2/4 > 1/N^2, X + Y + Z = 1\} \cup \{(X, Y, Z) : X^2 + Z^2 \leq \log^{-2} N, X^2 + Z^2/4 > 1/N^2, X + Y + Z = 1\}$ and $S_{\mathcal{D}} := \{(X, Y, Z) : (1, 0, 0), (1 - \frac{1}{N}, \frac{1}{N}, 0), (1 - \frac{1}{N}, 0, \frac{1}{N}), (\frac{1}{N}, 1 - \frac{1}{N}, 0), (0, 1 - \frac{1}{N}, \frac{1}{N}), (1 - \frac{2}{N}, 0, \frac{2}{N}), (0, 1 - \frac{2}{N}, \frac{2}{N}), (0, 1, 0)\}$ which will be used in later proof.

To make the definition more clear, we illustrate it using the picture below:



Now we are ready to state our main theorem:

Theorem 3.1.3. *If $G = K_N$, and $k = 2$, let $\tau := \inf\{t : (X^N, Y^N, Z^N)_t = (1,0,0)$ or $(0,1,0)\}$. Then, for any starting point $(X^N, Y^N, Z^N)_0 \in S_{\text{Inter}}$, $\tau = \log N + \log\log N + o(\log\log N)$ in probability as $N \to \infty$.*

3.2 Further conjectures

Here we list two conjectures which may worth further discussions.

- Theorem 3.1.3 claims that τ is $\log N + \log\log N + o(\log\log N)$, while we further conjecture the exact asymptotic form is possibly $\log N + \log\log N + O(1)$ instead;

- In theorem 3.1.3, we gave an asymptotic expression of the stopping time with initial configuration restricted to S_{Inter}. Further work may be focusing on the initials (1) near the boundary: $(x_0, y_0, z_0) \to \partial\mathcal{T}$, in such case we conjecture the time is of the form $\log N + O(1)$ without the $\log\log N$ term; (2) near the center: $|x_0 - y_0| \to 0$, where we conjecture the form $c^* \log N + o(\log N)$ with some constant $c^* > 1$.

Chapter 4

Proof of theorem 3.1.3

4.1 Basic Construction and Proof Outline

First, following [Fox16], we give definitions for drift and variance for general pure jump markov process.

Definition 4.1.1. Let $\{\mathbf{X}\}_t$ be a pure jump markov process defined on some compact set $K \subset \mathcal{R}^d$ adapted to filtration $\{\mathcal{F}_t : t \geq 0\}$. The total jump rate at time t is denoted as q_t which we assume is uniformly bounded. Then, there exists a probability measure ν_t on K s.t.

$$\lim_{h \to 0} \mathcal{P}(\mathbf{X}_t \text{ has a jump in } [t, t{+}h] \text{ by vector in } A|\mathcal{F}_t) \cdot h^{-1} = q_t \cdot \nu_t(A), \quad \forall t \geq 0, \forall A \in \mathcal{B}.$$

We define the jump for $\mathbf{X}$ to be $\Delta\mathbf{X}_t := \mathbf{X}_t - \mathbf{X}_{t^-}$ at time t, then define

$$\mu_t(\mathbf{X}) = q_t \int \Delta\mathbf{X}_t d\nu_t(\Delta\mathbf{X}_t)$$
$$\Sigma_t(\mathbf{X}) = q_t \int \Delta\mathbf{X}_t \cdot \Delta\mathbf{X}_t^T d\nu_t(\Delta\mathbf{X}_t),$$

to be the drift vector and covariance matrix for $\{\mathbf{X}\}_t$ respectively.

Now using the notations we mentioned in the previous section, we construct the formulas for both drift and coordinate variance for the $(X^N, Y^N, Z^N)_t$ process. Here the coordinate variance is just three variance terms in Σ_t^N, i.e. $(\Sigma_{11}^N, \Sigma_{22}^N, \Sigma_{33}^N)^T$ which we denote as σ_t^{N2} for convenience. For the rest covariance terms, we will see that $\Sigma_{12}^N = \Sigma_{21}^N$ is the only one needed in the later proofs, so we will derive the formula for it as well.

First, we note the jump rate for broadcasting one opinion for the whole system $\{X^N, Y^N, Z^N\}_t$ equals to N for any $t \geq 0$ since each person independently broadcasts an opinion at rate 1.

For the drift, according to the assumption of the rules for broadcasting, we have the following table:

	{A}	{B}	{A, B}
{A}	({A}, {A})	({A}, {A,B})	({A}, {A})
{B}	({B}, {A, B})	({B}, {B})	({B}, {B})
{A, B}	$\frac{1}{2}(\{A\}, \{A\})$	$\frac{1}{2}(\{A, B\}, \{A, B\})$	$\frac{1}{2}(\{A\}, \{A\})$
{A, B}	$\frac{1}{2}(\{A, B\}, \{A, B\})$	$\frac{1}{2}(\{B\}, \{B\})$	$\frac{1}{2}(\{B\}, \{B\})$

Table 4.1: Possible Changes of Opinions after Broadcasting

where the first column represents the opinion state for the person who is to broadcast his opinion, while the first row represents the opinion state for the person who is to receive opinion. Furthermore, the fourth row represents the resulting opinon set of both deliverer and receiver given the deliverer choosing to broadcast opinion A, while choosing to broadcast B in the fifth row. The $\frac{1}{2}$ factor in the last two rows means half-and-half probability for the resulting opinion set given the deliverer holding $\{A, B\}$.

We can transfer table 4.1 into the table of changes of $(X^N, Y^N, Z^N)_t$:

	{A}	{B}	{A, B}
{A}	$*$	(Y-, Z+)	(X+, Z-)
{B}	(X-, Z+)	$*$	(Y+, Z-)
{A, B}	(X+, Z-)	(Y-, Z+)	(X++, Z–)
{A, B}	(X-, Z+)	(Y+, Z-)	(Y++, Z–)

Table 4.2: Possible Changes of (X, Y, Z) after Broadcasting

where $*$ means no change, $\cdot + = \cdot + \frac{1}{N}$, $\cdot - = \cdot - \frac{1}{N}$, $\cdot ++ = \cdot + \frac{2}{N}$, and $\cdot -- = \cdot - \frac{2}{N}$.

Next, we compute the probability for all possibilities of the changes in (X^N, Y^N, Z^N), and it follows table 4.3 as below:

	{A}	{B}	{A, B}
{A}	$*$	$\frac{N}{N-1}XY$	$\frac{N}{N-1}XZ$
{B}	$\frac{N}{N-1}XY$	$*$	$\frac{N}{N-1}YZ$
{A, B}	$\frac{N}{2(N-1)}XZ$	$\frac{N}{2(N-1)}YZ$	$\frac{Z(NZ-1)}{2(N-1)}$
{A, B}	$\frac{N}{2(N-1)}XZ$	$\frac{N}{2(N-1)}YZ$	$\frac{Z(NZ-1)}{2(N-1)}$

Table 4.3: Probability for all Possible Changes

For example, the probability that deliverer holds $\{A\}$, and the receiver holds $\{B\}$ has probability $X \cdot \frac{NY}{N-1} = \frac{N}{N-1}XY$ given the current proportion vector being $\{X, Y, Z\}$. The rest probabilities can be similarly derived in table 4.3.

Therefore, according to definition 4.1.1, the drift for X_t^N equals to

$$N \cdot \frac{1}{N} \cdot (\frac{N}{N-1}(XZ - XY + Z^2) - \frac{Z}{N-1}) = \frac{N}{N-1}(XZ - XY + Z^2) - \frac{Z}{N-1},$$

which asymptotically equals to $(X^N Z^N - X^N Y^N + Z^{N2})_t$ as $N \to \infty$. Here in the first expression, the N term is the rate for broadcasting of the system, the $\frac{1}{N}$ term is the unit change of X_t^N, and the rest is the probability of different changes of X_t^N.

We can obtain formulas for drifts of Y_t^N and Z_t^N similarly, whence we have the drift vector μ_t^N for $(X^N, Y^N, Z^N)_t$ equals to:

$$(XZ - XY + Z^2, YZ - XY + Z^2, 2XY - XZ - YZ - 2Z^2) \cdot (1 + O(\frac{1}{N})). \tag{4.1.1}$$

Secondly, we can also compute the coordinate variance according to definition 4.1.1, table 4.2, and table 4.3. In particular, the variance for X_t^N equals to

$$N \cdot \frac{1}{N^2}((\frac{N}{N-1}(XZ+XY+XZ+2Z^2) - \frac{2Z}{N-1}) = \frac{1}{N-1}(XY+2XZ+2Z^2) - \frac{2Z}{N(N-1)},$$

where in the first expression the N term again represents the time rate for changing, the $\frac{1}{N^2}$ is the square of unit change of X_t^N, and the rest is the probability of different changes of X_t^N.

Similarly, we have formulas of coordinate variances for Y_t^N and Z_t^N, whence we have the coordinate variance vector σ_t^{N2} for $(X^N, Y^N, Z^N)_t$ equals to:

$$\frac{1}{N}(XY + 2XZ + 2Z^2, XY + 2YZ + 2Z^2, 2XY + 2YZ + 2XZ + Z^2) \cdot (1 + O(\frac{1}{N})). \tag{4.1.2}$$

Finally, we come to the covariance for X_t^N and Y_t^N. Since we observe that it is impossible for both X and Y to change values at the same time, so the covariance for them equals to 0, i.e. $\Sigma_{12}^N = 0$.

According to [Kur77], the pure jump markov process $(X^N, Y^N, Z^N)_t$ converges almost surely to $(X, Y, Z)_t$ as $N \to \infty$, where $(X, Y, Z)_t$ is a determinisitic process satisfying

$$d(X,Y,Z)_t = \mu_t(X,Y,Z)dt,$$
$$\mu_t(X,Y,Z) = (XZ - XY + Z^2, YZ - XY + Z^2, 2XY - XZ - YZ - 2Z^2)_t. \tag{4.1.3}$$

Furthermore, [Kur77] also showed in probability that

$$\sup_{t\le T} |(X^N, Y^N, Z^N)_t - (X,Y,Z)_t| \le (|(X^N, Y^N, Z^N)_0 - (X,Y,Z)_0| + \frac{C_1}{\sqrt{N}})e^{C_2 T} \tag{4.1.4}$$

for some constants $C_1, C_2 > 0$ independent of N and arbitrary $T > 0$, which we will be using repeatedly in chapter 4.2.

Therefore, our idea is to approximate this pure jump markov process by this deterministic model when the process is away from the two termination states $(1,0,0)$ and $(0,1,0)$, whereas instead of diffusion, we will use Poisson process and stochastic approximation via martingale methods to estimate the time when the process is approaching the termination points.

We prove theorem 3.1.3 by seperating the procedure into two parts. Note that X and Y are symmetric in terms of termination, WLOG, we assume the starting point is in the half region where $x_0 > y_0$ inside S_{Inter} for the rest sections.The outline of the proof is as follows:

I Given the starting point $\mathbf{v}_0 \in S_{\text{Inter}}$, denote $\tau_{\text{Ienter}} := \inf\{t : (X^N, Y^N, Z^N)_t \in S_{\log -dist}\}$, then $\tau_{\text{Inter}} = \log\log N + o(\log\log N)$ in probability;

II Given the starting point $\mathbf{v}_0 \in \partial S_{\log -dist}$, denote $\tau_{\log} := \inf\{t : (X^N, Y^N, Z^N)_t \in S_{\mathcal{D}}\}$, then $\tau_{\log} = \log N + o(\log\log N)$ in probability;

III It takes O(1) for any point $\mathbf{v}_0 \in S_{\mathcal{D}}$ to reach absorbing point.

In particular, we will prove I in chapter 4.2; prove II and III as well as completion of proof for the theorem in chapter 4.3.

4.2 Proof of I

First, before proving I, recall that we obtain the drift vector (4.1.3) for the corresponding determinisitc model in chapter 4.1, since here we focus ourselves to the case where $x_0 - y_0 \geq 2\epsilon_0$ which is straightforwardly depending on the process for $\{Y_t^N, Z_t^N\}$, basically we will be using both terms for Y_t^N and Z_t^N. Furthermore, we can rewrite them in terms of Y, Z using the equation $X + Y + Z = 1$. Namely,

$$\begin{aligned}\mu_t(Y, Z) &= (YZ - XY + Z^2, 2XY - XZ - YZ - 2Z^2)_t \\ &= (YZ - (1 - Y - Z)Y + Z^2, (1 - Y - Z)(2Y - Z) - YZ - 2Z^2)_t \\ &= (-Y + (Y + Z)^2, 2Y - Z - Y^2 - (Y + Z)^2)_t, \end{aligned} \tag{4.2.1}$$

we will be using drift in either form (4.1.3) or (4.2.1) when necessary.

In order to show $\tau_{\text{Inter}} = \log\log N + o(\log\log N)$ in probability, it is equivalent to show that: $\forall \delta > 0$, we have

$$(1 - \delta)\log\log N \leq \tau_{\text{Inter}} \leq (1 + \delta)\log\log N. \tag{4.2.2}$$

The outline to verify the above inequalities is as following:

1. $\exists L(\delta), \epsilon_1(\delta), \epsilon_2(\delta) > 0$ such that $Y_L^N \leq \frac{\delta}{2} Z_L^N$ and $\epsilon_1(\delta) \leq Y_L^N, Z_L^N \leq \epsilon_2(\delta)$ in probability;

2. $Y_t^N \leq \delta Z_t^N$ in probability for $L \leq t \leq \tau_{\text{Inter}}$, and we derive (4.2.2).

We verify the first statement using lemma 4.2.1-4.2.2.

Lemma 4.2.1. *Suppose we start at $(x_0, y_0, z_0) \in S_{\text{Inter}}$ satisfying $x_0 - y_0 \geq 2\epsilon_0$, then $\exists L(\epsilon^*)$ s.t.*

$$P(\epsilon^{**} \leq Y_L^N, Z_L^N \leq \epsilon^*) \to 1 \quad \text{as } N \to \infty,$$

for arbitrary $\epsilon^* > 0$ and $\epsilon^{**} > 0$ depending on ϵ^*.

Proof. It is sufficient to show for some $L(\epsilon^*)$

$$\epsilon^{**} \leq Y_L, Z_L \leq \epsilon^*$$

for the corresponding deterministic process due to (4.1.4) and $(X_0, Y_0, Z_0) = (x_0, y_0, z_0)$.

First, we define $U_t = X_t - Y_t$. Then $U_0 \geq 2\epsilon_0$.

According to (4.1.3), the drift of U_t equals to $Z_t U_t$ which is non-negative. Therefore, $U_t \geq 2\epsilon_0$ which indicates $Y_t \leq \frac{1}{2} - \epsilon_0$ for any t.

Second, let $V_t = (2+\epsilon_0)Y_t + Z_t$, according to (4.2.1), we have

$$\begin{aligned}
\mu(V_t) &= -(2+\epsilon_0)Y_t + (2+\epsilon_0)(Y_t+Z_t)^2 + 2Y_t - Z_t - Y_t^2 - (Y_t+Z_t)^2 \\
&= -\epsilon_0 Y_t + \epsilon_0 Y_t^2 - Z_t(1 - 2(1+\epsilon_0)Y_t - (1+\epsilon_0)Z_t) \\
&= -\epsilon_0 Y_t(1-Y_t) - Z_t(X_t - Y_t - \epsilon_0(1 - X_t + Y_t)) \\
&\leq -(\frac{1}{2}+\epsilon_0)\epsilon_0 Y_t - \epsilon_0 Z_t \\
&\leq -\frac{\epsilon_0}{4} V_t.
\end{aligned}$$

On the other hand, solving the following differential equation:

$$dV' = -\frac{\epsilon_0}{4}V' dt, \quad V_0' = V_0 = (2+\epsilon_0)y_0 + z_0$$

gives us

$$V_t' = V_0 e^{-\frac{\epsilon_0 t}{4}}.$$

Therefore, by letting $t_0 = 4\epsilon_0^{-1}(\log V_0 - \log \epsilon^*)$, we have

$$V_{t_0}' = \epsilon^*$$

which indicates

$$V_{t_0} \leq \epsilon^*.$$

Note that $V_{t_0} = (2+\epsilon_0)Y_{t_0} + Z_{t_0} \leq \epsilon^*$ implies both Y_{t_0} and Z_{t_0} are smaller than ϵ^*.

Third, we complete the proof by showing both Y_{t_0} and Z_{t_0} are greater than $\epsilon^{**} > 0$, where ϵ^{**} depends on ϵ^*.

In terms of (4.2.1), we have the following inequalities for $\mu(Y)$ and $\mu(Z)$:

$$\mu(Y) \geq -Y;$$

$$\mu(Z) \geq -2Z.$$

Therefore, the two following ODE's

$$dY^{'} = -Y^{'}dt, \quad Y_0^{'} = y_0;$$

$$dZ^{'} = -2Z^{'}dt, \quad Z_0^{'} = z_0$$

derives $Y_t^{'} = y_0 e^{-t}$ and $Z_t^{'} = z_0 e^{-2t}$.

If we let $\epsilon^{**} = 2\epsilon_0 e^{-2t_0}$, then

$$Y_{t_0} \geq \epsilon^{**},$$

$$Z_{t_0} \geq \epsilon^{**},$$

here we use the assumption that $Y_0, Z_0 \geq 2\epsilon_0$.

In sum, let $L(\epsilon^*) = t_0$, we conclude

$$\epsilon^{**} \leq Y_L, Z_L \leq \epsilon^*$$

for arbitrary $\epsilon^* > 0$, whence

$$P(\epsilon^{**} \leq Y_L^N, Z_L^N \leq \epsilon^*) \to 1 \quad \text{as } N \to \infty,$$

we are done. □

Lemma 4.2.2. *Suppose we start at* (x_0, y_0, z_0) *such that* $\epsilon^{**}(\delta) \leq y_0, z_0 \leq \epsilon^*(\delta)$, *then* $\exists L, T(\delta), L \leq T(\delta)$ *s.t.*

$$Y_L^N \leq \frac{\delta}{2} Z_L^N \quad \textit{in probability},$$

where $\epsilon^* = \frac{\delta}{16(1+2/\delta)^2}$ *and* ϵ^{**} *depends on* ϵ^* *in terms of lemma 4.2.1.*

Proof. As similar to lemma 4.2.1, it is sufficient to show

$$Y_L \le \frac{\delta}{2} Z_L \quad \text{for some } L \le T(\delta)$$

for the corresponding deterministic model because of (4.1.4) and $T(\delta)$ not depending on N.

First, from the proof in lemma 4.2.1, we see $V_t = (2+\epsilon_0)Y_t + Z_t$ is non-increasing. Therefore, $V_t \le V_0 \le 4\epsilon^*$ which implies $Y_t, Z_t \le 4\epsilon^*$ for all t.

Second, we define $\tau := \inf\{t : \frac{\delta}{2} Z_t - Y_t \ge 0\}$. If $y_0 \le \frac{\delta}{2} z_0$, then we are done. Otherwise, we have $y_0 > \frac{\delta}{2} z_0$. Therefore, $\forall t < \tau$, we have

$$\begin{aligned}
\mu(Y) &\le -Y + (1+2/\delta)^2 Y^2 \\
&\le -(1 - 4(1+2/\delta)^2 \epsilon^*) Y; \\
\mu(Z) &\ge \delta Z - Z - (\delta^2/4 + (1+\delta/2)^2) Z^2 \\
&\ge -(1 - \delta + 4(\delta^2/4 + (1+\delta/2)^2)\epsilon^*) Z.
\end{aligned}$$

Furthermore, if we plug in $\epsilon^* = \frac{\delta}{16(1+2/\delta)^2}$, we deduce

$$\begin{aligned}
\mu(Y) &\le -(1-\delta/4) Y; \\
\mu(Z) &\ge -(1-\delta/2) Z.
\end{aligned}$$

Therefore, we have the following inequalities:

$$\begin{aligned}
Y_t &\le y_0 e^{-(1-\delta/4)t}; \\
Z_t &\ge z_0 e^{-(1-\delta/2)t}.
\end{aligned}$$

Let $t_0 = \frac{4}{\delta}\log\frac{\epsilon^*}{\delta\epsilon^{**}}$, then we have $Y_{t_0} \le \frac{\delta}{2}Z_{t_0}$.

Finally, let $T(\delta) = t_0$, we have $\tau \le T(\delta)$. Thus, we are done with the proof. $\square$

Next, we verify the second statement using lemma 4.2.3-4.2.4. Before we go ahead, from lemma 4.2.2, we see that $\exists L$ independent of N (in fact, $L \le \frac{4}{\delta}\log\frac{\epsilon^*}{\delta\epsilon^{**}}$) such that $Y_L \le \frac{\delta}{2}Z_L$, and both Y_L and Z_L are bounded between two positive numbers independent of N (in fact, the one candidate for upper bound is $4\epsilon^* = \frac{\delta}{4(1+2/\delta)^2}$). We denote ϵ_3 and ϵ_4 as the lower and upper bounds for Y_L and Z_L respectively for the later proof.

Lemma 4.2.3. *Suppose we start at (x_0, y_0, z_0) such that $y_0 \le \frac{\delta}{2}z_0$ and $\epsilon_3 \le y_0, z_0 \le \epsilon_4$, then*

$$Y_t^N \le \delta Z_t^N \quad \forall t < 2\log\log N \quad \textit{in probability.}$$

Proof. First, let $T = 2\log\log N$ in (4.1.4), we have

$$\begin{aligned}\sup_{t\le 2\log\log N} |(X^N, Y^N, Z^N)_t - (X,Y,Z)_t| &\le \frac{C_1}{\sqrt{N}}e^{C_2\log\log N}\\ &= \frac{C_1\log^{C_2}N}{\sqrt{N}}\\ &\ll \log^{-1}N \text{ in probability.} \qquad (4.2.3)\end{aligned}$$

We will be using (4.2.3) later in this proof.

Second, from lemma 4.2.2, we see $2Y_t + Z_t \le 4\epsilon_4$ for all t. let $W_t = \delta Z_t - Y_t$, then $W_0 = \delta z_0 - y_0 \ge y_0 \ge \epsilon_3$.

We claim: $\exists\theta(\delta) > 0$ s.t.

$$\mu(W_t) \ge -(1+\theta)W_t,$$

which is equivalent to

$$
\begin{aligned}
& -\delta Z+(1+2\delta)Y-\delta Y^2-(1+\delta)(Y+Z)^2 \geq -(1+\theta)(\delta Z-Y) \\
\Leftrightarrow \quad & \theta\delta Z+(2\delta-\theta)Y-\delta Y^2-(1+\delta)(Y+Z)^2>0 \\
\Leftarrow \quad & (\theta\delta-4(1+\delta)\epsilon_4)Z+(2\delta-\theta-4\delta\epsilon_4-4(1+\delta)\epsilon_4)Y>0 \\
\Leftarrow \quad & \theta\delta>4(1+\delta)\epsilon_4, \quad 2\delta>\theta+4\delta\epsilon_4+4(1+\delta)\epsilon_4 \\
\Leftrightarrow \quad & \frac{1+\delta}{(1+2/\delta)^2}<2\delta-\frac{\delta+2\delta^2}{(1+2/\delta)^2} \\
\Leftrightarrow \quad & 1+2\delta+2\delta^2<8/\delta+8+2\delta,
\end{aligned}
$$

which is true for small δ. Here, we use $\epsilon_4=\frac{\delta}{4(1+2/\delta)^2}$ for the above inequalities. Therefore, we have $W_t \geq W_0 e^{-(1+\theta)t}$ for all t.

Finally, if we restrict $t\in[0,2\log\log N]$, applying (4.2.3) provides us:

$$
\begin{aligned}
\inf_{t\leq 2\log\log N}\delta Z_t^N-Y_t^N &= \inf_{t\leq 2\log\log N}\delta(Z_t^N-Z_t)-(Y_t^N-Y_t)+\delta Z_t-Y_t \\
&\geq \inf_{t\leq 2\log\log N}\delta Z_t-Y_t-\sup_{t\leq 2\log\log N}|(X^N,Y^N,Z^N)_t-(X,Y,Z)_t| \\
&\geq W_0 e^{-2(1+\theta)\log\log N}-\frac{C_1\log^{C_2}N}{\sqrt{N}} \\
&\geq \epsilon_3\log^{-2(1+\theta)}N-\frac{C_1\log^{C_2}N}{\sqrt{N}} \\
&\geq 0 \quad \text{as } N\to\infty,
\end{aligned}
$$

which completes the proof of this lemma. □

Finally, we complete this section by proving:

Lemma 4.2.4 (I). *Suppose we start at* (x_0,y_0,z_0) *such that* $y_0\leq\frac{\delta}{2}z_0$ *and* $\epsilon_3\leq$

$y_0, z_0 \leq \epsilon_4$, *then*

$$(1-\delta)\log\log N \leq \tau_{\text{Inter}} \leq (1+\delta)\log\log N \textit{ in probability.}$$

Proof. First, due to (4.2.3), it is sufficient to show

$$(1-\delta)\log\log N \leq \tau^* \leq (1+\delta)\log\log N$$

where $\tau^* = \inf\{t : Y_t^2 + Z_t^2 \leq \log^{-2} N\}$ for the corresponding deterministic model.

According to lemma 4.2.3, we know that $Y_t \leq \delta Z_t$ and $Y_t, Z_t \leq 4\epsilon_4$ for $t \leq 2\log\log N$. We restrict the time interval to be $[0, 2\log\log N]$ in the rest of the proof.

Second, we consider the process $U_t = 3Y_t + Z_t$. The drift has the following bounds:

$$\begin{aligned}
\mu(U) &= -U + 2Y - Y^2 + 2(Y+Z)^2 \\
&\geq -U; \\
\mu(U) &= -U + 2Y - Y^2 + 2(Y+Z)^2 \\
&\leq -U + \frac{2}{3+1/\delta}U + 16\epsilon_4 U \\
&\leq -(1-\frac{\delta}{2})U
\end{aligned}$$

for sufficiently small δ.

Thus we derive the bounds for U_t:

$$U_0 e^{-t} \leq U_t \leq U_0 e^{-(1-\frac{\delta}{2})t}$$

which indicates $(1-\delta)\log\log N \le \tau^* \le (1+\delta)\log\log N$ because

$$U_0 \log^{-1-\delta/2+\delta^2/2} N \ll \log^{-1} N \ll U_0 \log^{-1+\delta} N$$

Therefore, we have

$$(1-\delta)\log\log N \le \tau_{\text{Inter}} \le (1+\delta)\log\log N \quad \text{in probability,}$$

we are done. $\square$

To sum up, lemma 4.2.1-4.2.4 implies the total amount of time from S_{Inter} to reaching $S_{\log -dist}$ is $O(1)+O(1)+\log\log N+o(\log\log N) = \log\log N+o(\log\log N)$ which completes the proof of I.

4.3 Proof of II and III

In this chapter, the main idea is to construct some martingale arguments to prove II. In particular, lemma 4.3.4, 5, 7, 8 will deduce some geometric properties for some process we construct; lemma 4.3.9 will study the original $(X^N, Y^N, Z^N)_t$ process through some random walk estimation when the process is within $S_{\log -dist}$; lemma 4.3.11 illustrates the core estimation using all previous lemmas. Eventually, we complete our proof for theorem 3.1.3 at the end of this chapter.

Before stating the lemmas, we need the following definitions:

Definition 4.3.1. We define the 2-dimensional flow as follows:

$$F(y,z) = (-y, 2y - z);$$

$\Phi(y,z,t)$ satisfies:

$$\frac{\partial}{\partial t}\Phi(y,z,t) = F(\Phi(y,z,t)),$$

$$\Phi(y,z,0) = (y,z),$$

where $F(y,z)$ is the linear aproximation for (4.2.1).

Furthermore, we introduce the hitting time function

$$\Psi(y,z) = \inf\{t : \Phi(y,z,t) \in \partial E\},$$

where $E = \{(y,z) : y^2 + \frac{z^2}{4} \leq \frac{1}{N^2}\}$. And we let

$$\mathcal{H}(y,z) = \begin{bmatrix} \partial_{yy}\Psi & \partial_{yz}\Psi \\ \partial_{yz}\Psi & \partial_{zz}\Psi \end{bmatrix}$$

denote the Hessian matrix for Ψ.

By definition, we have

$$\tau_{\log} = \inf\{t : (Y_t^N, Z_t^N) \in E\}.$$

Remark 4.3.2. For all the following notations and calculations, we use the conventions for rows versus columns as follows:

Vector valued functions such as F and Φ are written as row vectors;

Coordinates of a derivative occupy columns. Thus, the gradient of a scalar function is a column vector. However, the derivative of gradient, i.e., Hessian is regarded as a matrix, not as a tensor with two types of columns.

Moreover we will use $\langle u, v\rangle$ for the product of a row vector u and a column vector v, so that the notation makes clear that the quantity is a scalar. The Hessian of a scalar function such as G is denoted

$$\mathcal{H}(G) := \left(\frac{\partial^2 G}{\partial x_i \partial x_j}\right)_{i,j}.$$

Definition 4.3.3. We define

$$W_t := \Psi(Y_t^N, Z_t^N) + t,$$

where $\{Y_t^N, Z_t^N\}$ is the stochastic process of opinion broadcasting model at time t.

Lemma 4.3.4. *$\forall \mathbf{v}_0 = (y_0, z_0) \in S_{\log -dist}$, we have:*

$$\max\{\frac{1}{6}, \log \frac{|\mathbf{v}_0|N}{2}\} \leq \Psi(\mathbf{v}_0) \leq \log N + 1.$$

Furthermore, if $\mathbf{v}_0 = (y_0, z_0) \in \partial S_{\log -dist}$, i.e. $|\mathbf{v}_0| = \log^{-1} N$, then

$$\log N - o(\log \log N) \leq \Psi(\mathbf{v}_0).$$

Proof. First, we can solve the flow Φ in terms of F as follows:

$$y_t = y_0 e^{-t};$$

$$z_t = (2y_0 t + z_0)e^{-t}.$$

Then,

$$\Psi(\mathbf{v}_0) = \inf_t \{y_t^2 + \frac{z_t^2}{4} = \frac{1}{N^2}\}.$$

Notice that $\mathbf{v}_0$ is outside of E and if y_0, z_0 takes only integer values when multiplied by N, we have $|\mathbf{v}_0| \geq \frac{\sqrt{2}}{N}$. Therefore, let $t_1 = \frac{1}{6}$, we have

$$\begin{aligned} y_{t_1}^2 + \frac{z_{t_1}^2}{4} &= (y_0^2 + \frac{1}{4}(\frac{1}{3}y_0 + z_0)^2)e^{-1/3} \\ &\geq \frac{13}{9e^{1/3}} \cdot \frac{1}{N^2} \\ &\geq \frac{1}{N^2}, \end{aligned}$$

whence $\Psi(\mathbf{v}_0) \geq \frac{1}{6}$.

Furthermore, if $|\mathbf{v}_0| \geq \frac{2}{N}$, let $t_1 = \log(|\mathbf{v}_0|N/2)$, then $t_1 \geq 0$ and

$$\begin{aligned} y_{t_1}^2 + \frac{z_{t_1}^2}{4} &= (y_0^2 t_1^2 + y_0 z_0 t_1 + (y_0^2 + \frac{z_0^2}{4}))e^{-2t_1} \\ &= \frac{4}{|\mathbf{v}_0|^2 N^2} \cdot (y_0^2 t_1^2 + y_0 z_0 t_1 + y_0^2 + \frac{z_0^2}{4}) \\ &\geq \frac{4y_0^2 + z_0^2}{|\mathbf{v}_0|^2 N^2} \\ &\geq \frac{1}{N^2}, \end{aligned}$$

whence $\Psi(\mathbf{v}_0) \geq \log(|\mathbf{v}_0|N/2)$.

Let $t_2 = \log N + 1$, then

$$\begin{aligned} y_{t_2}^2 + \frac{z_{t_2}^2}{4} &= (y_0^2 t_2^2 + y_0 z_0 t_2 + (y_0^2 + \frac{z_0^2}{4}))e^{-2t_2} \\ &= \frac{1}{e^2 N^2} \cdot (y_0^2(\log N + 1)^2 + y_0 z_0(\log N + 1) + y_0^2 + \frac{z_0^2}{4}) \\ &\leq \frac{1}{e^2 N^2} \cdot (1 + \frac{4}{\log N}) \\ &\leq \frac{1}{N^2}, \end{aligned}$$

provided that $|\mathbf{v}_0| \leq \log^{-1} N$. Therefore,

$$\Psi(\mathbf{v}_0) \leq \log N + 1.$$

Finally, if $|\mathbf{v}_0| = \log^{-1} N$, then according to chapter 4.2, we notice that

$$\begin{aligned} y_0 &\geq e^{-\log\log N - \omega} \\ &= \frac{1}{\log N \cdot e^{\omega}}, \end{aligned}$$

where $\omega = o(\log\log N), e^{\omega} = o(\log N)$.

Thus if we let $t_3 = \log N - 2\omega$, we have

$$\begin{aligned} y_{t_3}^2 + \frac{z_{t_3}^2}{4} &= (y_0^2 t_3^2 + y_0 z_0 t_3 + (y_0^2 + \frac{z_0^2}{4}))e^{-2t_3} \\ &\geq y_0^2 t_3^2 e^{-2t_3} \\ &= \frac{1}{N^2} \cdot \frac{(\log N - 2\omega)^2 \cdot e^{4\omega}}{\log^2 N \cdot e^{2\omega}} \\ &\geq \frac{1}{N^2}, \end{aligned}$$

whence $\Psi(\mathbf{v}_0) \geq \log N - 2\omega = \log N - o(\log\log N)$, we are done. $\square$

Lemma 4.3.5. *We can write both the gradient and Hessian of Ψ in terms of Φ, F, and B:*

$$\nabla\Psi(\mathbf{v}_0) = -\frac{d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\nabla B(\mathbf{v}_1)}{\langle F(\mathbf{v}_1), \nabla B(\mathbf{v}_1)\rangle}; \tag{4.3.1}$$

$$\begin{aligned} \mathcal{H}(\Psi)(\mathbf{v}_0) &= \nabla\Psi_1(\mathbf{v}_0) \cdot \mathcal{H}(\Phi_1)(\mathbf{v}_0) + \nabla\Psi_2(\mathbf{v}_0) \cdot \mathcal{H}(\Phi_2)(\mathbf{v}_0) \\ &+ d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\frac{dF\nabla B\nabla B^T}{\langle F, \nabla B\rangle^2}(\mathbf{v}_1) \cdot d\Phi(\cdot, t_0)(\mathbf{v}_0) \\ &- d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\left[I - \frac{\nabla BF}{\langle F, \nabla B\rangle}\right]\frac{\mathcal{H}(B)}{\langle F, \nabla B\rangle}(\mathbf{v}_1) \cdot d\Phi(\cdot, t_0)(\mathbf{v}_0) \\ &+ d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\frac{\nabla B\nabla B^T}{\langle F, \nabla B\rangle^2}\left[I - \frac{\nabla BF}{\langle F, \nabla B\rangle}\right]dF(\mathbf{v}_1) \cdot d\Phi(\cdot, t_0)(\mathbf{v}_0), \end{aligned} \tag{4.3.2}$$

where $t_0 = \Psi(\mathbf{v}_0), \mathbf{v}_1 = \Phi(\mathbf{v}_0, t_0)$; $\Phi = (\Phi_1, \Phi_2)$, and $\nabla\Psi = (\nabla\Psi_1, \nabla\Psi_2)^T$; $B(y, z) = y^2 + \frac{z^2}{4} - \frac{1}{N^2}$ indicating $B(\mathbf{v}_1) = 0$ by definition.

Proof. We first tackle the special case where $t_0 = 0$, that is where the initial point $\mathbf{v}_0$ already satisfies $B(\mathbf{v}_0) = 0$. In this case the time t map $\Phi(\cdot, t)$ is the identity, therefore $d\Phi = I$ and $\mathcal{H}(\Phi) = 0$. Thus (4.3.1) and (4.3.2) reduce to

$$\nabla\Psi = -\frac{\nabla B}{\langle F, \nabla B\rangle} \tag{4.3.3}$$

$$\begin{aligned}\mathcal{H}(\Psi) = &\frac{dF\nabla B\nabla B^T}{\langle F, \nabla B\rangle^2} - \left[I - \frac{\nabla BF}{\langle F, \nabla B\rangle}\right]\frac{\mathcal{H}(B)}{\langle F, \nabla B\rangle} \\ &+ \frac{\nabla B\nabla B^T}{\langle F, \nabla B\rangle}\left[I - \frac{\nabla BF}{\langle F, \nabla B\rangle}\right]\frac{dF}{\langle F, \nabla B\rangle},\end{aligned} \tag{4.3.4}$$

where we omit the point at which the functions are valued since in this special case $\mathbf{v}_0 = \mathbf{v}_1$, so no confusions should be there.

In order to verify these two equations above, we consider calculating the gradient and Hessian for $B(\Phi(\mathbf{v}, \eta(\mathbf{v})))$ where η is any C^2 function with $\mathbf{v}$ satisfying $\eta(\mathbf{v}) = 0$. Applying basic calculus and chain rules, we have

$$\nabla B(\Phi(\mathbf{v}, \eta(\mathbf{v}))) = \nabla(\mathbf{v}) + \nabla\eta(\mathbf{v}) \cdot \langle F(\mathbf{v}), \nabla B(\mathbf{v})\rangle; \tag{4.3.5}$$

$$\begin{aligned}\mathcal{H}(B(\Phi(\mathbf{v}, \eta(\mathbf{v})))) = &\, dF\nabla\eta\nabla B(\mathbf{v}) + [I + \nabla\eta F]\,\mathcal{H}(B)(\mathbf{v}) \\ &+ \mathcal{H}(\eta)F\nabla B(\mathbf{v}) + \nabla\eta\nabla B^T\,[I + \nabla\eta F]\,dF(\mathbf{v}).\end{aligned} \tag{4.3.6}$$

Now notice the function Ψ satisfies $B(\Phi(\mathbf{v}, \Psi(\mathbf{v}))) \equiv 0$, therefore the gradient and Hessian vanish. Also, $\Psi(\mathbf{v}_0) = 0$ since $B(\mathbf{v}_0) = 0$. Letting $\eta = \Psi$ in (4.3.5) and (4.3.6), and setting the right-hand sides equal to zero, we may solve for $\nabla\Psi$ and $\mathcal{H}(\Psi)$ to obtain (4.3.3) and (4.3.4) respectively.

The general case relies on the following explicit chain rule for any composition of maps:

Remark 4.3.6. Let $\Phi : \mathcal{R}^m \to \mathcal{R}^n$ and $\Psi : \mathcal{R}^n \to \mathcal{R}$ be any twice continuously differentiable maps. Then

$$\nabla[\Psi \circ \Phi] = d\Phi \nabla \Psi,$$
$$\mathcal{H}[\Psi \circ \Phi] = \sum_{i=1}^{n} \nabla \Psi_i \mathcal{H}(\Phi_i) + d\Phi^T \mathcal{H}(\Psi) d\Phi,$$

where $\Phi = (\Phi_1, \Phi_2, \cdots, \Phi_n), \nabla\Psi = (\nabla\Psi_1, \nabla\Psi_2, \cdots, \nabla\Psi_n)^T$.

With this in mind, observe that for fixed s, the quantity $\Psi \circ \Phi(\cdot, s)$ differs from Ψ by the constant s. Letting $s = t_0 := \Psi(\mathbf{v}_0)$, and recalling the notation $\mathbf{v}_1 = \Phi(\mathbf{v}_0, t_0)$, we see that $\nabla\Psi(\mathbf{v}_0) = \nabla[\Psi \circ \Phi(\cdot, t_0)]$ at $\mathbf{v}_0$ and $\mathcal{H}(\Psi)(\mathbf{v}_0) = \mathcal{H}[\Psi \circ \Phi(\cdot, t_0)]$ at $\mathbf{v}_0$. In both cases, the derivatives of Ψ are evaluated only where B vanishes. Therefore, we may use remark 4.3.6 along with (4.3.3) and (4.3.4) to obtain (4.3.1) and (4.3.2) respectively. □

Lemma 4.3.7.

$$\|\nabla\Psi(\mathbf{v}_0)\| \leq C_1 \cdot \log |\mathbf{v}_0| N \cdot \|\mathbf{v}_0\|^{-1}$$

for $\mathbf{v}_0 = (y_0, z_0) \in S_{\log - dist}$.

Proof. We follow the same notations we used in lemma 4.3.5. Then lemma 4.3.5 implies:

$$\nabla\Psi(\mathbf{v}_0) = -\frac{d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\nabla B(\mathbf{v}_1)}{\langle F(\mathbf{v}_1), \nabla B(\mathbf{v}_1)\rangle}.$$

Therefore,

$$\begin{aligned}
\|\nabla\Psi(\mathbf{v})_0\| &\leq \frac{\sqrt{10}\|d\Phi(\cdot, t_0)(\mathbf{v}_0)\| \cdot \|\nabla B(\mathbf{v}_1)\|}{\|F(\mathbf{v}_1)\| \cdot \|\nabla B(\mathbf{v}_1)\| \cdot |\cos\theta|} \\
&= \frac{\sqrt{10}\|d\Phi(\cdot, t_0)(\mathbf{v}_0)\|}{\|F(\mathbf{v}_1)\| \cdot |\cos\theta|},
\end{aligned}$$

where the matrix norm is defined to be $\|\cdot\|_\infty$, the maximum absolute value of all elements in the matrix; θ is the angle between the vectors F and ∇B.

In particular, notice that

$$\begin{aligned}
\langle F(\mathbf{v}_1), \nabla B(\mathbf{v}_1)\rangle &= \langle(-y_1, 2y_1 - z_1), (2y_1, \frac{z_1}{2})\rangle \\
&= -2y_1^2 + y_1 z_1 - \frac{z_1^2}{2} \\
&= -\frac{1}{2}(y_1 - z_1)^2 - \frac{3}{2}y_1^2 \\
&< 0,
\end{aligned}$$

hence $\cos\theta$ is uniformly up-bounded by some negative constant, i.e. $|\cos\theta| > M_1 > 0$.

Next, since $\mathbf{v}_1$ satisfies $B(\mathbf{v}_1) = 0$,

$$\|F(\mathbf{v}_1)\| > M_2 \cdot \frac{1}{N},$$

for some constant $M_2 > 0$. Moreover, we have

$$\|d\Phi(\cdot, t_0)(\mathbf{v}_0)\| = 2t_0 \cdot e^{-t_0}.$$

According to lemma 4.3.4, if $|\mathbf{v}_0| \geq \frac{2e}{N}$, then

$$\begin{aligned}2t_0 \cdot e^{-t_0} &\leq 4\log\frac{|\mathbf{v}_0|N}{2} \Big/ |\mathbf{v}_0|N \\ &\leq M_3 \cdot \frac{\log|\mathbf{v}_0|N}{|\mathbf{v}_0|N};\end{aligned}$$

otherwise if $\frac{\sqrt{2}}{N} \leq |\mathbf{v}_0| < \frac{2e}{N}$,

$$\begin{aligned}2t_0 \cdot e^{-t_0} &\leq \frac{2}{e} \\ &\leq \frac{4\sqrt{2}}{e\log 2} \cdot \frac{\log|\mathbf{v}_0|N}{|\mathbf{v}_0|N}.\end{aligned}$$

Summing up the above inequalities, we proved the lemma. □

Lemma 4.3.8.

$$\|\mathcal{H}(\Psi)(\mathbf{v}_0)\| \leq C_2 \cdot \log^2|\mathbf{v}_0|N \cdot \|\mathbf{v}_0\|^{-2},$$

for $\mathbf{v}_0 = (y_0, z_0) \in S_{\log-dist}$.

Proof. Again lemma 4.3.5 states:

$$\begin{aligned}\mathcal{H}(\Psi)(\mathbf{v}_0) = &\nabla\Psi_1(\mathbf{v}_0) \cdot \mathcal{H}(\Phi_1)(\mathbf{v}_0) + \nabla\Psi_2(\mathbf{v}_0) \cdot \mathcal{H}(\Phi_2)(\mathbf{v}_0) \\ &+ d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\frac{dF\nabla B\nabla B^T}{\langle F, \nabla B\rangle^2}(\mathbf{v}_1) \cdot d\Phi(\cdot, t_0)(\mathbf{v}_0) \\ &- d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\left[I - \frac{\nabla BF}{\langle F, \nabla B\rangle}\right]\frac{\mathcal{H}(B)}{\langle F, \nabla B\rangle}(\mathbf{v}_1) \cdot d\Phi(\cdot, t_0)(\mathbf{v}_0) \\ &+ d\Phi(\cdot, t_0)^T(\mathbf{v}_0)\frac{\nabla B\nabla B^T}{\langle F, \nabla B\rangle^2}\left[I - \frac{\nabla BF}{\langle F, \nabla B\rangle}\right]dF(\mathbf{v}_1) \cdot d\Phi(\cdot, t_0)(\mathbf{v}_0),\end{aligned}$$

where $\Phi = (\Phi_1, \Phi_2)$, and $\nabla\Psi = (\nabla\Psi_1, \nabla\Psi_2)^T$. We can further write $\nabla\Psi$ in terms of Φ, F, B, but there is no necesssity to do that here.

In order to estimate the norm for $\mathcal{H}(\mathbf{v}_0)$, it is sufficient to estimate the norm of each term in the above expression, and the dominant term will determine the entire one. Before estimation, we do some calculation for preparation:

$$\begin{aligned}
\Phi_1(\mathbf{v}_0, t_0) &= y_0 e^{-t_0} \\
\Rightarrow \mathcal{H}(\Phi_1)(\mathbf{v}_0) &= \mathbf{0} \\
\Phi_2(\mathbf{v}_0, t_0) &= (2y_0 t - z_0) e^{-t_0} \\
\Rightarrow \mathcal{H}(\Phi_2)(\mathbf{v}_0) &= \mathbf{0},
\end{aligned}$$

thus the first two terms disappeared.

$$\begin{aligned}
\|d\Phi(\cdot, t_0)(\mathbf{v}_0)\| &\leq \frac{M_1 \log |\mathbf{v}_0| N}{|\mathbf{v}_0| N} \\
\|I - \frac{\nabla BF}{\langle F, \nabla B \rangle}\| &\leq M_3 \\
\|F(\mathbf{v}_1)\| &> M_2 \cdot \frac{1}{N} \\
\|dF(\mathbf{v}_1)\| &= 2 \\
\|\nabla B(\mathbf{v}_1)\| &> M_4 \cdot \frac{1}{N} \\
\|\mathcal{H}(B)(\mathbf{v}_1)\| &= 2,
\end{aligned}$$

for some constants $M_1, M_2, M_3, M_4 > 0$.

Therefore, we see that each of the rest term has the same order. In particular, each term is up-bounded by

$$M_0 \cdot \frac{\log^2 |\mathbf{v}_0| N}{|\mathbf{v}_0|^2 N^2} \cdot N^2,$$

whence the norm of $\mathcal{H}(\Psi)(\mathbf{v}_0)$ is bounded by

$$C_2 \cdot \log^2 |\mathbf{v}_0| N \cdot \|\mathbf{v}_0\|^{-2}.$$

$\square$

Now that we have completed the lemmas on geometric properties for process $\{W_t\}$, we move on to study some properties for the original process $(X^N, Y^N, Z^N)_t$ within $S_{\log -dist}$ (lemma 4.3.9). Eventually we will combine them to deduce lemma 4.3.11.

Lemma 4.3.9. *Suppose the process is inside $S_{\log -dist}$, define the set of time stripes*

$$\mathcal{T}_i = \{s : 2^{-i-1} \leq \|\mathbf{v}_s\| < 2^{-i}\},$$

where $\lceil \log_2 \log N \rceil \leq i \leq \lfloor \log_2 N \rfloor - 1$, $\|\mathbf{v}_s\|^2 = Y_s^{N2} + Z_s^{N2}$.

Then we have $\max |\mathcal{T}_i|$ is $O(1)$ in probability.

Proof. We consider the process $U_t^N = 3Y_t^N + Z_t^N$, which has negative drift which is upper bounded by $U_t^N/6$. U_t^N is a pure jump markov process with jump rate N. Let $N(t)$ be the number of jumps between $[0, t]$, then $N(t) \sim Pois(Nt)$. Furthermore, we define V_n^N be the discrete random walk embedded in U_t^N. Namely, $U_t^N = V_{N(t)}^N, \forall t$. Before stating the proof in detail, note that we have the relation between U_t^N and $\|\mathbf{v}_t\|$:

$$\|\mathbf{v}_t\|^2 \leq U_t^{N2} \leq 12\|\mathbf{v}_t\|^2,$$

which indicates that instead of studying $\{\mathcal{T}_i\}$, we can equivalently estimate the probability for $\max|\mathcal{T}_i'|$ greater than some big constant, where

$$\mathcal{T}_i' = \{s : 2^{-i-1} \le U_t^N < 2^{-i}\}, \quad \lceil \log_2 \log N \rceil \le i \le \lfloor \log_2 N \rfloor - 1.$$

Therefore, WLOG, we prove $\max|\mathcal{T}_i'|$ is $O(1)$ in probability. Let $\tau_i = \inf\{t : U_t^N \le 2^{-i}\}, d_i = \tau_{i+1} - \tau_i, \lceil \log_2 \log N \rceil \le i \le \lfloor \log_2 N \rfloor - 1$. The idea is to show

- $\max d_i$ is $O(1)$ in probability;
- $\max_i \sup_{\tau_i \le t \le \tau_{i+1}} U_t^N / 2^{-i}$ is $O(1)$ in probability.

Then, the result follows since each $\mathcal{T}_i$ can cover at most constant number of time intervals $[\tau_i, \tau_{i+1}]$ and all d_i's are constant. We seperate the proof into four parts.

First, we show the following inequality:

$$P(\tau_V > CN) \le e^{-\alpha CNA}, \quad \text{for some constant } \alpha > 0 \tag{4.3.7}$$

where $\tau_V = \inf\{k : V_k^N \le \frac{A}{2}\}$, and $V_0^N = A$ for A between N^{-1} and $\log^{-1} N$.

Define $V_k^N = V_0^N + \sum_{i=1}^k M_i$, where M_i corresponds to each jump.

According to table 4.2 and table 4.3 in chapter 4.1, we know that M_i has the following distribtution given current position $(X, Y, Z)_{i-1}$:

- $-\frac{2}{N}$ with probability $(XY + \frac{YZ}{2} + \frac{Z^2}{2}) \cdot (1 + O(\frac{1}{N})) := p_{-2}$;
- $-\frac{1}{N}$ with probability $\frac{3}{2}XZ \cdot (1 + O(\frac{1}{N})) := p_{-1}$;
- $\frac{1}{N}$ with probability $(XY + \frac{1}{2}XZ) \cdot (1 + O(\frac{1}{N})) := p_1$;

- $\frac{2}{N}$ with probability $\frac{3}{2}YZ \cdot (1 + O(\frac{1}{N})) := p_2$;

- $\frac{4}{N}$ with probability $\frac{1}{2}Z^2 \cdot (1 + O(\frac{1}{N})) := p_4$;

- 0 otherwise $:= p_0$.

We claim: $\exists \theta > 0$ such that:

$$E(e^{\theta M_i} | V_{i-1}^N > \frac{A}{2}) \leq 1 - \frac{A}{M'} \quad \text{for some constant } M^{'}$$

uniformly with respect to i.

By definition, we have

$$\begin{aligned}
E(e^{\theta M_i} | V_{i-1}^N > \frac{A}{2}) &= p_{-2}e^{-2\theta/N} + p_{-1}e^{-\theta/N} + p_0 + p_1 e^{\theta/N} + p_2 e^{2\theta/N} + p_4 e^{4\theta/N} \\
&= 1 + \theta E(M_i | V_{i-1}^N > \frac{A}{2}) + O(\theta^2 E(M_i^2 | V_{i-1}^N > \frac{A}{2})) \\
&\leq 1 - \frac{\theta}{6n} V_{i-1}^N + O(\frac{\theta^2}{N^2} V_{i-1}^N) \\
&\leq 1 - \frac{\theta}{12N} A + O(\frac{\theta^2}{N^2} A).
\end{aligned}$$

Here, the first inequality follows from the fact that U_t^N's drift is upper bounded by $-\frac{1}{6}U_t^N$, and the second inequality follows from the condition $V_{i-1}^N > \frac{A}{2}$.

Therefore, if we let $\theta = \frac{N}{M}$ for some large constant M, then since $M_i \sim \frac{1}{N}$, $\theta M_i \sim \frac{1}{M}$ is small. Furthermore,

$$E(e^{\theta M_i} | V_{i-1}^N) \leq 1 - \frac{A}{12M} + O(\frac{A}{M^2}) \leq 1 - \frac{A}{M'},$$

for some constants $M^{'} > 0$, we are done for the claim.

Therefore, if we pick such θ, we have that

$$
\begin{aligned}
P(\tau_V > CN) &= P(V_k^N > \frac{A}{2}, \forall k \leq CN) \\
&= P(V_{CN}^N \cdot \prod_{k=0}^{CN-1} \mathbf{1}_{\{V_k^N > A/2\}} > \frac{A}{2}) \\
&= P(\sum_{i=1}^{CN} M_i \cdot \prod_{k=0}^{CN-1} \mathbf{1}_{\{V_k^N > A/2\}} > -\frac{A}{2}) \\
&= P(e^{\theta \sum_{i=1}^{CN} M_i} \cdot \prod_{k=0}^{CN-1} \mathbf{1}_{\{V_k^N > A/2\}} > e^{-\theta \frac{A}{2}}) \\
&\leq E(e^{\theta \sum_{i=1}^{CN} M_i} \cdot \prod_{k=0}^{CN-1} \mathbf{1}_{\{V_k^N > A/2\}}) \cdot e^{\theta \frac{A}{2}} \\
&= E(e^{\theta \sum_{i=1}^{CN-1} M_i} \cdot \prod_{k=0}^{CN-2} \mathbf{1}_{\{V_k^N > A/2\}} E(e^{\theta M_{CN}} | V_{CN-1}^N > \frac{A}{2})) \cdot e^{\frac{1}{2M}} \\
&\leq E(e^{\theta \sum_{i=1}^{CN-1} M_i} \cdot \prod_{k=0}^{CN-2} \mathbf{1}_{\{V_k^N > A/2\}}) \cdot (1 - \frac{A}{M'}) \cdot e^{\frac{1}{2M}} \\
&\leq \cdots\cdots \\
&\leq (1 - \frac{A}{M'})^{CN} \cdot e^{\frac{1}{2M}} \\
&= K e^{-\frac{CNA}{M'}},
\end{aligned}
$$

which completes our proof of (4.3.7)

Second, we convert V_i^N to U_t^N according to the relation $U_t^N = V_{N(t)}^N$. According to Chernoff bound argument, for any Possion random variable $X \sim Pois(\nu)$, we have the bound for tail probability:

$$
P(X \leq x) \leq \frac{e^{-\nu}(e\nu)^x}{x^x}, \quad \text{for } x < \nu.
$$

Let $\nu = Nt, x = \nu/2$, we obtain for U_t^N,

$$P(\#\text{ jumps } < Nt/2 \text{ between } [0,t]) \leq e^{-\nu} \cdot (2e)^x$$
$$= \left(\frac{2}{e}\right)^x.$$

Therefore, $P(\#\text{ jumps } \geq Nt/2) \geq 1 - \left(\frac{2}{e}\right)^{Nt/2}$.

Combining the first two parts, we have

$$\begin{aligned} P(d_i < 2C) &= P(\inf_{\tau_i < t < \tau_i + 2C} U_t^N \leq \frac{A}{2}) \\ &\geq P(\tau_V \leq CN | \#\text{ jumps } \geq CN) \cdot P(\#\text{ jumps } \geq CN) \\ &\geq (1 - e^{-\alpha CNA}) \cdot \left(1 - \left(\frac{2}{e}\right)^{CN}\right), \end{aligned}$$

where $A = 2^{-i}$, $i = \lceil \log_2 \log N \rceil, \ldots \lfloor \log_2 N \rfloor - 1$.

Third, we show

$$P(\max_{i: N^{-1} \leq 2^{-i} \leq \log^{-1} N} d_i < C) \geq e^{-2e^{-2\alpha C}}.$$

We do the calculation as follows:

$$\begin{aligned} P(\max_{i: N^{-1} \leq 2^{-i} \leq \log^{-1} N} d_i < C) &= \prod_i P(d_i < C) \\ &\geq \prod_i \left[\left(1 - \left(\frac{2}{e}\right)^{CN/2}\right) \cdot (1 - e^{-\alpha CN/2^i})\right]. \end{aligned} \tag{4.3.8}$$

Taking logarithm of (4.3.8), we obtain

$$\begin{aligned} \sum_i \left[\log\left(1 - \left(\frac{2}{e}\right)^{CN/2}\right) + \log(1 - e^{-\alpha CN/2^i})\right] &\sim -\sum_i \left[\left(\frac{2}{e}\right)^{CN/2} - e^{-\alpha CN/2^i}\right] \\ &\geq -\log N \cdot \left(\frac{2}{e}\right)^{CN/2} - \frac{e^{-2\alpha C}}{1 - e^{-2\alpha C}} \\ &\geq -2e^{-2\alpha C} \quad \text{for large constant } C, \end{aligned}$$

which completes the proof for the third part.

Finally, we show

$$P(\sup_{\tau_i \le t \le \tau_{i+1}} U_t^N \le (CM+1)\cdot 2^{-i}, \forall i = \lceil \log_2 \log N \rceil, \dots \lfloor \log_2 N \rfloor - 1) \ge e^{-2e^{-2C}}.$$

Notice that from the first part of the proof we can pick $\theta = \frac{N}{M}$ such that $e^{\theta U_t^N}$ is a non-negative supermartingale due to U_t^N has negative drift, we apply Doob's maximal inequality to $e^{\theta U_t^N}$ within each time interval $[\tau_i, \tau_{i+1}]$, we obtain

$$\begin{aligned}
P(\sup_{\tau_i \le t \le \tau_{i+1}} U_t^N > (CM+1)\cdot 2^{-i}) &= P(\sup_{\tau_i \le t \le \tau_{i+1}} e^{\theta U_t^N} > e^{\theta(CM+1)\cdot 2^{-i}}) \\
&\le \frac{E(e^{\theta U_{\tau_i}^N})}{e^{\theta(CM+1)\cdot 2^{-i}}} \\
&\le e^{-\theta CM 2^{-i}} \\
&= e^{-CN2^{-i}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&P(\sup_{\tau_i \le t \le \tau_{i+1}} U_t^N \le (CM+1)\cdot 2^{-i}, \forall i = \lceil \log_2 \log N \rceil, \dots \lfloor \log_2 N \rfloor - 1) \\
=&\prod_i P(\sup_{\tau_i \le t \le \tau_{i+1}} U_t^N \le (CM+1)\cdot 2^{-i}) \\
\ge&\prod_i (1 - e^{-CN2^{-i}}) \qquad (4.3.9)
\end{aligned}$$

Taking logarithm of (4.3.9), we have

$$\begin{aligned}
\sum_{\lceil \log_2 \log N \rceil \le i \le \lfloor \log_2 N \rfloor - 1} \log(1 - e^{-CN2^{-i}}) &\sim -\sum_i e^{-CN2^{-i}} \\
&\ge -\frac{e^{-2C}}{1 - e^{-2C}} \\
&\ge -2e^{-2C} \quad \text{for larget constant C,}
\end{aligned}$$

which completes the proof for the last part.

To sum up, we reach our conclusion that $\max |\mathcal{T}_i|$ is $O(1)$ in probability. $\square$

Remark 4.3.10 (III). When the process enters $S_{\mathcal{D}}$, we verify that it takes O(1) time to reach $(1, 0, 0)$.

Inside $S_{\mathcal{D}}$, the possible value set for U_0^N is $\{\frac{1}{N}, \frac{2}{N}, \frac{3}{N}\}$. Following exactly the same arguemts in lemma 4.3.9, we see that, in probability, it takes O(1) time for the pure jump process to reach $\frac{U_0^N}{4} \leq \frac{1}{N}$. Since each jump step is a multiple of $1/N$, we see that when the process goes below $\frac{U_0^N}{4}$, it reaches termination.

It is time to add up all the above 5 lemmas to verify statement II as below:

Lemma 4.3.11 (II)**.** $W_{\mathcal{T}_{\log}} = \log N + O(1)$ *in probability given the starting point* $(x_0, y_0, z_0) \in \partial S_{\log -dist}$.

Proof. First, for convenience, we denote $\mathbf{v}_t = (Y_t^N, Z_t^N)$. Note that $\mathbf{v}_t$ is a pure jump markov process, by definition 4.1.1, we can write

$$\mathbf{v}_t(\omega) = \mathbf{v}_0(\omega) + \sum_{\text{jump happens at } s(\omega) < t} \Delta(\mathbf{v}_s(\omega)),$$

then for $\Psi(\cdot)$,

$$\Psi(\mathbf{v}_t(\omega)) = \Psi(\mathbf{v}_0(\omega)) + \sum_{\text{jump happens at } s(\omega) < t} \Delta\Psi(\mathbf{v}_s(\omega)),$$

where $\Delta\Psi(\mathbf{v}_s(\omega)) = \Psi(\mathbf{v}_s(\omega)) - \Psi(\mathbf{v}_{s^-}(\omega))$. In the below, we will use notations without ω for convenience.

We write $W_t = M_t + A_t$, where $A_t = \int_0^t N\cdot\left[\int (\Psi(\mathbf{v}_s+\Delta(\mathbf{v}_s)) - \Psi(\mathbf{v}_s)) d\nu_s(\Delta(\mathbf{v}_s))\right] ds + t$. Recall the N term in the formula for A_t is the transition rate for the original process. Notice that

$$
\begin{aligned}
E(M_t|\mathcal{F}_s) &= E(W_t - A_t|\mathcal{F}_s) \\
&= E(\Psi(\mathbf{v}_t) - \int_0^t N\cdot\left[\int (\Psi(\mathbf{v}_r+\Delta(\mathbf{v}_r)) - \Psi(\mathbf{v}_r)) d\nu_r(\Delta(\mathbf{v}_r))\right] dr|\mathcal{F}_s) \\
&= E(\Psi(\mathbf{v}_t)|\mathcal{F}_s) - N\left(\int_s^t + \int_0^s\right)\left[E\int (\Psi(\mathbf{v}_r+\Delta(\mathbf{v}_r)) - \Psi(\mathbf{v}_r)) d\nu_r(\Delta(\mathbf{v}_r))|\mathcal{F}_s\right] dr \\
&= \Psi(\mathbf{v}_s) - N\int_0^s\left[\int (\Psi(\mathbf{v}_r+\Delta(\mathbf{v}_r)) - \Psi(\mathbf{v}_r)) d\nu_r(\Delta(\mathbf{v}_r))\right] dr \\
&= W_s - A_s \\
&= M_s.
\end{aligned}
$$

So $\{M_t\}$ is a local martingale.

Next, we calculate and estimate A_t by seperating the process in terms of which strip the current position is located. In particular, we consider the time stripes as defined in lemma 4.3.9:

$$
\mathcal{T}_i = \{s : 2^{-i} \le \|\mathbf{v}_s\| < 2^{-i+1}\}, \quad N^{-1} \le 2^{-i} \le \log^{-1} N.
$$

Then,

$$
\begin{aligned}
A_t &= \int_0^t N\cdot\left[\int (\Psi(\mathbf{v}_s+\Delta(\mathbf{v}_s)) - \Psi(\mathbf{v}_s)) d\nu_s(\Delta(\mathbf{v}_s))\right] ds + t \\
&= \sum_i N\int_{\mathcal{T}_i}\left[\int (\Psi(\mathbf{v}_s+\Delta(\mathbf{v}_s)) - \Psi(\mathbf{v}_s)) d\nu_s(\Delta(\mathbf{v}_s))\right] ds + |\mathcal{T}_i|.
\end{aligned}
$$

Within $\mathcal{T}_i$, according to Taylor expansion, we have:

$$
\begin{aligned}
& N\int_{\mathcal{T}_i}\left[\int(\Psi(\mathbf{v}_s+\Delta(\mathbf{v}_s))-\Psi(\mathbf{v}_s))d\nu_s(\Delta(\mathbf{v}_s))\right]ds+|\mathcal{T}_i| \\
=& N\int_{\mathcal{T}_i}\left[\int(\nabla\Psi(\mathbf{v}_s)\cdot\Delta(\mathbf{v}_s)+\mathcal{R}_s)d\nu_s(\Delta(\mathbf{v}_s))\right]ds+\int_{\mathcal{T}_i}ds \\
=& \int_{\mathcal{T}_i}(\nabla\Psi(\mathbf{v}_s)\cdot\mu_s(\mathbf{v}_s)+1)ds+N\int_{\mathcal{T}_i}\left[\int\mathcal{R}_s d\nu_s(\Delta(\mathbf{v}_s))\right]ds \\
=& \int_{\mathcal{T}_i}(\nabla\Psi(\mathbf{v}_s)\cdot F(\mathbf{v}_s)+1+\nabla\Psi(\mathbf{v}_s)\cdot(\mu_s-F)(\mathbf{v}_s))ds+N\int_{\mathcal{T}_i}\left[\int\mathcal{R}_s d\nu_s(\Delta(\mathbf{v}_s))\right]ds,
\end{aligned}
\tag{4.3.10}
$$

where by (4.2.1),

$$
(\mu_s-F)(\mathbf{v}_s)=((Y+Z)^2,-Y^2-(Y+Z)^2)_s\cdot(1+O(\frac{1}{N}));
$$

and

$$
\mathcal{R}_s:=\Psi(\mathbf{v}_s+\Delta(\mathbf{v}_s))-\Psi(\mathbf{v}_s)-\nabla\Psi(\mathbf{v}_s)\cdot\Delta(\mathbf{v}_s)
$$

is the residual part in the Taylor expansion beyond the linear term.

We claim: $\nabla\Psi(\mathbf{v}_s)\cdot F(\mathbf{v}_s)+1=0, \quad \forall s.$

By lemma 4.3.5, we have

$$
\begin{aligned}
\nabla\Psi(\mathbf{v}_s) &= -\frac{d\Phi(\cdot,t_s)^T(\mathbf{v}_s)\nabla B(\mathbf{v}_1)}{\langle F(\mathbf{v}_1),\nabla B(\mathbf{v}_1)\rangle} \\
&= -\begin{bmatrix} e^{-t_s} & 2t_s e^{-t_s} \\ 0 & e^{-t_s}\end{bmatrix}(2y_1,\frac{z_1}{2})^t/(-y_1,2y_1-z_1)\cdot(2y_1,\frac{z_1}{2})^t \\
&= (2y_1e^{-t_s}+z_1t_se^{-t_s},\frac{z_1e^{-t_s}}{2})^t/(2y_1^2-y_1z_1+\frac{z_1^2}{2}).
\end{aligned}
$$

Recall that $y_1=y_se^{-t_s}$ and $z_1=(2y_st_s+z_s)e^{-t_s}$, so we have $y_s=e^{t_s}y_1, z_s=$

$z_1 e^{t_s} - 2y_1 t_s e^{t_s}$. Therefore,

$$
\begin{aligned}
\nabla\Psi(\mathbf{v}_s)\cdot F(\mathbf{v}_s) &= (-y_s, 2y_s - z_s)\cdot(2y_1 e^{-t_s} + z_1 t_s e^{-t_s}, \frac{z_1 e^{-t_s}}{2})^t/(2y_1^2 - y_1 z_1 + \frac{z_1^2}{2}) \\
&= \frac{-2y_1 y_s e^{-t_s} - y_s t_s e^{-t_s} z_1 + y_s e^{-t_s} z_1 - z_s e^{-t_s} z_1/2}{2y_1^2 - y_1 z_1 + z_1^2/2} \\
&= \frac{-2y_1^2 + y_1 z_1 - z_1^2/2}{2y_1^2 - y_1 z_1 + z_1^2/2} \\
&= -1,
\end{aligned}
$$

we proved the claim. Thus (4.3.10) becomes

$$
\int_{\mathcal{T}_i} \nabla\Psi(\mathbf{v}_s)\cdot(\mu_s - F)(\mathbf{v}_s)ds + N\int_{\mathcal{T}_i}\left[\int \mathcal{R}_s d\nu_s(\Delta(\mathbf{v}_s))\right]ds. \tag{4.3.11}
$$

For the first part of (4.3.11), lemma 4.3.7 indicates:

$$
\begin{aligned}
\int_{\mathcal{T}_i} \nabla\Psi(\mathbf{v}_s)\cdot(\mu_s - F)(\mathbf{v}_s)ds &\le \sup_{t\in\mathcal{T}_i} \|\nabla\Psi(\mathbf{v}_t)\|\cdot|(\mu_t - F)(\mathbf{v}_s)|\cdot|\mathcal{T}_i| \\
&\le \sup_{t\in\mathcal{T}_i} 5C_1\cdot\log|\mathbf{v}_t|N\cdot\|\mathbf{v}_t\|^{-1}\cdot\|\mathbf{v}_t\|^2\cdot|\mathcal{T}_i| \\
&\le 5C_1\log N\cdot 2^{-i+1}\cdot|\mathcal{T}_i|.
\end{aligned} \tag{4.3.12}
$$

For the second part of (4.3.11), according to Taylor inequality, we have

$$
\begin{aligned}
& N\int_{\mathcal{T}_i}\left[\int \mathcal{R}_s d\nu_s(\Delta(\mathbf{v}_s))\right]ds \\
&\le \frac{1}{2}\sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} \|\mathcal{H}(\Psi)(\mathbf{v}_t)\|\cdot N\cdot\int_{\mathcal{T}_i} E((\Delta Y_s^N)^2 + 2\Delta Y_s^N\Delta Z_s^N + (\Delta Z_s^N)^2|\mathcal{F}_s)ds \\
&\le \sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} \|\mathcal{H}(\Psi)(\mathbf{v}_t)\|\cdot N\cdot\int_{\mathcal{T}_i} E((\Delta Y_s^N)^2 + (\Delta Z_s^N)^2|\mathcal{F}_s)ds \\
&= \sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} \|\mathcal{H}(\Psi)(\mathbf{v}_t)\|\int_{\mathcal{T}_i}\sigma^2(Y_s^N) + \sigma^2(Z_s^N)ds,
\end{aligned}
$$

where the right-hand-side of the first inequality is due to the fact that the norm of the position after one jump from $\mathbf{v}_t$ such that $2^{-i} \leq \|\mathbf{v}_t\| < 2^{-i+1}$ will be within $[2^{-i-1}, 2^{-i+2})$. The reason is that one jump can at most change the norm by $\frac{1}{N}$ which is smaller than any of $2^{-i}, i = \lfloor \log_2 \log N \rfloor, \cdots \lfloor \log_2 N \rfloor$.

Recall from (4.1.2), we have:

$$\sigma_s^{N2}(Y,Z) = \frac{1}{N}(XY + 2YZ + 2Z^2, 2XY + 2YZ + 2XZ + Z^2)_s + O(\frac{1}{N^2}).$$

If we write it in terms of Y and Z using the equation $X + Y + Z = 1$, we obtain:

$$\sigma_s^{N2}(Y,Z) = \frac{1}{N}(Y - Y^2 + YZ + 2Z^2, 2Y + 2Z - 2Y^2 - 2YZ - Z^2)_s + O(\frac{1}{N^2}),$$

which implies

$$\sigma^2(Y_s^N) + \sigma^2(Z_s^N) \leq \frac{1}{N} \cdot 4 \cdot 2^{-i+2}$$

within $\mathcal{T}_i$; and lemma 4.3.8 indicates:

$$\sup_{t \in \mathcal{T}_{i-1}, \mathcal{T}_i, \mathcal{T}_{i+1}} \|\mathcal{H}(\Psi)(\mathbf{v}_t)\| \leq \sup_{t \in \mathcal{T}_{i-1}, \mathcal{T}_i, \mathcal{T}_{i+1}} C_2 \cdot \log^2 |\mathbf{v}_t| N \cdot \|\mathbf{v}_t\|^{-2} \leq C_2 \cdot \log^2 2^{-i+2} N \cdot 2^{2i+2}.$$

Therefore,

$$N \int_{\mathcal{T}_i} \left[\int \mathcal{R}_s d\nu_s(\Delta(\mathbf{v}_s)) \right] ds \leq 64 C_2 \cdot \frac{2^i}{N} \cdot \log^2 2^{-i+2} N \cdot |\mathcal{T}_i|. \tag{4.3.13}$$

By lemma 4.3.9, we know that the time for $\{Y_t^N, Z_t^N\}$ staying in $\mathcal{T}_i$ for any i is $O(1)$ in probability, and notice that we only need $\mathcal{T}_i$'s for $2^{-i} \geq \frac{1}{N}$ because we only consider the process W_t until time $\tau_{\log}$. So adding up (4.3.12) and (4.3.13) we obtain an upper bound for A_t as follows:

$$
\begin{aligned}
A_t &\leq C \sum_{i:N^{-1}\leq 2^{-i}\leq \log^{-1} N} 2^{-i}\log N + \frac{2^i}{N}\cdot \log^2 2^{-i+2}N \\
&\leq 2C + \frac{4C}{N}\sum_{i:2^{-i}\geq \frac{2}{N}} 2^i \cdot (\log N - i\log 2)^2
\end{aligned}
\tag{4.3.14}
$$

where C is some constant, and $0 \leq t \leq \tau_{\log}$.

Denote $k = \log_2 N$, then $\log 2 \cdot k = \log N$. We claim that

$$
\sum_{i=0}^{k-1} 2^i(\log N - i\log 2)^2
$$

is up-bounded by $C^{'} \cdot N$, for some constant $C^{'}$, which is sufficient to verify (4.3.14) being bounded by some constant.

We have

$$
\begin{aligned}
&\sum_{i=0}^{k-1} 2^i(\log N - i\log 2)^2 \\
&= \sum_{i=0}^{k-1} 2^i \cdot \log^2 N - 2\log 2\log N\sum_{i=0}^{k-1} i\cdot 2^i + \log^2 2\sum_{i=0}^{k-1} i^2\cdot 2^i \\
&= (2^k-1)\log^2 N - 2\log 2\log N((k-2)2^k+2) + \log^2 2((k^2-4k+6)2^k-6) \\
&= -\log^2 N - 4\log 2\log N + 6\log^2 2\cdot 2^k - 6\log^2 2 \\
&= 6\log^2 2\cdot N - \log^2 N - 4\log 2\log N - 6\log^2 2 \\
&\leq C^{'}N.
\end{aligned}
\tag{4.3.15}
$$

Thus, we have shown $A_t = O(1)$. Therefore, $A_{\tau_{\log}} = O(1)$ in probablity.

Next we claim that the quadratic variation for M_t is also bounded by $O(1)$.

According to [Fox16], we have

$$\langle M\rangle_t = \int_0^t \sigma_s^2(M)ds, \quad \text{and} \quad \sigma_t^2(M) = N\cdot E[\Delta_t^2(M)|\mathcal{F}_t].$$

We calculate $\sigma_t^2(M)$ first, for t in some $\mathcal{T}_i$. Notice that process A_t contains no jumps, we have the following:

$$\begin{aligned}
\sigma_t^2(M) &= N\cdot E[\Delta_t^2(M)|\mathcal{F}_t] \\
&= N\cdot E[\Delta_t^2(\Psi)|\mathcal{F}_t] \\
&\le \sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} \|\nabla\Psi(\mathbf{v}_t)\|^2\cdot N\cdot E[(\Delta Y_t^N + \Delta Z_t^N)^2|\mathcal{F}_t] \\
&\le \sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} \|\nabla\Psi(\mathbf{v}_t)\|^2\cdot 2N\cdot E[(\Delta Y_t^N)^2 + (\Delta Z_t^N)^2|\mathcal{F}_t] \\
&= \sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} \|\nabla\Psi(\mathbf{v}_t)\|^2\cdot 2\cdot(\sigma^2(Y_s^N) + \sigma^2(Z_s^N)) \\
&\le \sup_{t\in\mathcal{T}_{i-1},\mathcal{T}_i,\mathcal{T}_{i+1}} C_1\cdot\log^2|\mathbf{v}_t|N\cdot\|\mathbf{v}_t\|^{-2}\cdot 2\cdot\frac{1}{N}\cdot 4\cdot 2^{-i+2} \\
&\le 128C_1\cdot\log^2 2^{-i+2}N\cdot\frac{2^i}{N},
\end{aligned} \tag{4.3.16}$$

where the first inequality follows the same reason as the estimation for residual $\mathcal{R}_t$ in previous proof.

Next, using (4.3.16), we obtain the upper bound on $\langle M\rangle_t$:

$$\begin{aligned}\langle M\rangle_t &= \int_0^t \sigma_s^2(M)ds \\ &= \sum_{i:2^{-i}\geq\frac{1}{N}} \int_{\mathcal{T}_i} \sigma_s^2(M)ds \\ &\leq \frac{C}{N}\sum_{i:2^{-i}\geq\frac{1}{N}} 2^i\cdot\log^2 2^{-i+2}N \\ &= O(1), \end{aligned} \tag{4.3.17}$$

where the last equation comes from (4.3.15), using which we verified $A_t = O(1)$ as well.

Finally, let $t = \tau_{\log}$, since $\langle M\rangle_{\tau_{\log}} = O(1)$ implies bounded variance, i.e. $E(M_{\tau_{\log}} - M_0)^2 < C_*$ for some constant $C_* > 0$, applying Doob's maximal inequality, we obtain

$$P(|M_{\tau_{\log}} - M_0| > L) \leq P(\sup_{0\leq t\leq\tau_{\log}} |M_t - M_0| > L) \leq \frac{C_*}{L^2}. \tag{4.3.18}$$

Therefore, $\forall\epsilon > 0$, let $L = \sqrt{\frac{C_*}{\epsilon}}$, (4.3.18) implies

$$P(|M_{\tau_{\log}} - M_0| > L) \leq \epsilon, \tag{4.3.19}$$

namely, $|M_{\tau_{\log}} - M_0| = O(1)$ in probability. Thus we have

$$W_{\tau_{\log}} = M_{\tau_{\log}} + A_{\tau_{\log}} = M_0 + O(1) = \log N + o(\log\log N) \quad \text{in probability,}$$

where the first equality is by definition; the second one is from (4.3.15) and (4.3.19); the third one is from lemma 4.3.4.

□

So far, remark 4.3.10 plus lemma 4.3.11 completes the arguments for II and III.

Finally, combining with chapter 4.2, we have proved I, II, III, whence we complete our proof for theorem 3.1.3.

Bibliography

[Ben99] M. Benaïm. *Dynamics of stochastic approximation algorithms*. Seminaires de Probabilités XXXIII, volume 1709 of Lecture notes in mathematics, 1999.

[BH96] M. Benaïm and M. Hirsch. *Asymptotic pseudotrajectories and chain-recurrent flows, with applications*. J. Dynam. Differential Equations, 1996.

[EP23] F. Eggenberger and G. Pólya. *Über die Statistik vorketter vorgänge*. Zeit. Angew. Math. Mech., 1923.

[Fox16] Eric Foxall. *Stochastic calculus and sample path estimation for jump processes*. Math.PR, 2016.

[HLS80] B. Hill, D. Lane, and W. Sudderth. *A strong law for some generalized urn processes*. Ann. Probab., 1980.

[Kur77] T.G. Kurtz. *Strong approximation theorems for density dependent markov chains*. Elsevier, vol. 6(3), 1977.

[Pem88] R. Pemantle. *Random processes with reinforcement*. Doctoral Dissertation. M.I.T., 1988.

[RM51] H. Robbins and S. Monro. *A stochastic approximation method*. Ann. Math. Statist., 22:400407, 1951.

[RW00] L.C.G.Rogers and David Williams. *Diffusions, Markov Processes, and Martingales, Vol 2: Ito Calculus, P269*. Cambridge, 2000.